Comparison of the National Emergency Department Overcrowding Scale and the Emergency Department Work Index for Quantifying Emergency Department Crowding

Steven J. Weiss, MD, Amy A. Ernst, MD, Todd G. Nick, PhD

Abstract

Background: Emergency department (ED) crowding is just beginning to be quantified. The only two scales presently available are the National Emergency Department Overcrowding Scale (NEDOCS) and the Emergency Department Work Index (EDWIN).

Objectives: To assess the value of the NEDOCS and the EDWIN in predicting overcrowding. The hypothesis of this study was that the NEDOCS and the EDWIN would be equally sensitive and specific for overcrowding.

Methods: The NEDOCS, the EDWIN, and an overcrowding measure (OV) were determined every two hours for a ten-day period in December 2004. The NEDOCS is a statistically derived calculation, and the EDWIN is a formula-based calculation. The overcrowding measure is a composite of physician and charge nurse expert opinion on the degree of overcrowding as measured on a 100-mm visual analogue scale (VAS). The primary outcome, overcrowding, was based on the dichotomized OV VAS score at the midpoint of 50 mm ($\geq$50, overcrowded; <50, not overcrowded). The area under the receiver operator characteristic curve (AUC) and an index of adequacy (relative prognostic content) of each measure, on the basis of the likelihood ratio chi-square statistic, were computed to evaluate the performance of NEDOCS and EDWIN.

Results: There were 130 completed sampling times over ten days. The OV indicated that the ED was overcrowded 62% of the time. The AUC for the NEDOCS was 0.83 (95% CI = 0.75 to 0.90), and the AUC for the EDWIN was 0.80 (95% CI = 0.73 to 0.88). The NEDOCS score accounts for 97% of the prognostic information provided by combining all variables used in each model into one combined model. The EDWIN score accounts for only 86% (χ^2 test for difference, p = 0.02).

Conclusions: Both scales had high AUCs, correlated well with each other, and showed good discrimination for predicting ED overcrowding. This establishes construct validity for these scales as measures of overcrowding. Which scale is used in an ED is dependent on which set of data is most readily available, with the favored scale being the NEDOCS.

ACADEMIC EMERGENCY MEDICINE 2006; 13:513–518 © 2006 by the Society for Academic Emergency Medicine

Keywords: overcrowding, scales

From the Department of Emergency Medicine, University of New Mexico (SJW, AAE), Albuquerque, NM; and Department of Pediatrics, Cincinnati Children's Hospital Medical Center and Cincinnati College of Medicine (TGN), Cincinnati, OH.

Received October 6, 2005; revision received December 4, 2005; accepted December 4, 2005.

Address for correspondence and reprints: Steven Weiss, MD, University of New Mexico, MSC10 5560, 1 University of New Mexico, Albuquerque, NM 87131-0001. Fax: 505-272-6503; e-mail: sweiss52@aol.com.

In 2004, Hwang and Concato stated, "Although emergency department (ED) overcrowding has been a topic of frequent investigation, current definitions of the problem are often implicit or focus on factors outside of the ED itself. A more consistent approach to defining ED overcrowding would help to clarify the distinctions between causes, characteristics, and outcomes."[1] In the past there has been no standardization and no generalizable definition of overcrowding. Even the American College of Emergency Physicians'

doi: 10.1197/j.aem.2005.12.009
ISSN 1069-6563
PII ISSN 1069-6563583

definition is difficult to apply across different EDs with differing problems.

Emergency department overcrowding is only beginning to be studied in ways that allow for standardized quantification of the problem.[2–5] Methods of quantification have been based on different theoretical constructs, but all use emergency providers' opinions as an outcome variable. The two primary scales available at present for quantifying overcrowding are the NEDOCS[5] and the EDWIN model.[3] Although each was developed by using different methodology, both attempt to model the same outcome variable of real-time expert opinion on ED overcrowding.

The purposes of this study were to determine the validity of the NEDOCS and EDWIN in comparison to ED expert opinion on overcrowding, to determine whether there was a similar construct validity represented by both scales, and to evaluate the two scales in comparison to each other as measures of overcrowding.

METHODS

Study Design

This was a prospective study of the assessment mechanisms for ED overcrowding. The study did not involve any patient contact, and all identifiers were removed from the information as it was obtained. The institutional review board approved this study as exempt from informed consent requirements.

Study Setting and Population

The study was performed in an inner-city, Level 3 trauma center with an ED patient census of >60,000 per year.

Study Protocol

An overcrowding measure (OV) was calculated on the basis of the method used by the NEDOCS investigators. Our ED always has two physicians and one charge nurse covering the unit at all times. At sampling times, both ED physicians and the ED charge nurse were asked to rate the degree of overcrowding on a 100-mm visual analogue scale (VAS). Results were combined into a composite outcome score by taking the average of the physician and nurse scores. The primary outcome, overcrowding, was based on the OV VAS score being above or below the midpoint of the VAS (overcrowded, $\geq$50 mm; not overcrowded, <50 mm).

The NEDOCS Model. The first model evaluated was the NEDOCS. The criteria for NEDOCS variables were the following: 1) represented a snapshot of the ED, 2) represented an aspect of ED patient management (for example, triage, treatment, and disposition), 3) was readily available, 4) was definable such that results were reproducible between observers, and 5) was consistently defined between institutions. The variable of patient acuity, the definition of which varies greatly between institutions, was not used. The NEDOCS was created in a stepwise fashion, leading to a reduced model of five variables. The reduced model of overcrowding includes the following items: 1) ED patients (indexed to ED beds), 2) number of ventilators in use in the ED, 3) longest admit time, 4) waiting room time for the last patient called to a bed, and 5) indexed admits in the ED (indexed to hospital beds). These items were entered into a developed algorithm and yielded a score between 1 and 200, with less than 100 considered not overcrowded and more than or equal to 100 considered overcrowded. Within this spectrum, six categories exist, from *not busy* to *dangerously overcrowded*. The NEDOCS was used exactly as described in earlier publications.[5,6]

The EDWIN Model. EDWIN is defined as

$$\frac{\Sigma n_i t_i}{N_a(B_T - B_A)},$$

where n_i = number of patients in the ED in the triage category *i*, t_i = triage category, N_a = the number of attending physicians on duty, B_T = the number of treatment bays, and B_A = the number of admitted patients in the ED.[3] The triage system used was the Emergency Severity Index (ESI), a five-level instrument that has high interobserver agreement and is associated with resource use and hospitalization rates.[7]

Sampling Methods. This study calculated both the NEDOCS and the EDWIN scores every two hours for a ten-day period. All values for the EDWIN model were available for download from our computerized triage system. For the NEDOCS, the computerized triage system easily presented the number of patients and the number of admissions. For the calculation of ED wait and admit times, computerized system snapshots were downloaded every 5 minutes, which allowed us to determine these times within a 5-minute time frame. We obtained the number of respirator patients in the ED from an attending physician. The outcome variable was obtained by asking the ED attendings and the charge nurse to rate independently the level of ED crowding. The VAS was a 100-mm line that used six-point Likert levels similar to those used to validate the NEDOCS. The Likert scale corresponding to levels of overcrowding was shown to all participants, next to the VAS. Neither of the two scales had been derived previously or validated in this particular ED.

All data necessary for the ED overcrowding scales were readily available. Results were calculated for each scale on the basis of published information.

Data Analysis

Descriptive statistics were used to characterize the sampling times. Pearson correlation coefficients were calculated for testing associations among the two overcrowding scales and the composite OV. Our primary outcome was *overcrowding, yes/no*. Overcrowding was set to *yes* if the OV VAS score was at least 50 mm. Otherwise it was set to *no*. Receiver operator characteristic (ROC) curves were used to compare the association between sensitivity and specificity of an overcrowding scale for various cut points, thus allowing the determination of an optimal cut point. To determine the predictive ability of the overcrowding scales, three logistic regression models were used to predict overcrowding from NEDOCS alone, EDWIN alone, and both scales. Predictive discrimination of the scales was determined by using the C-statistic, which is a generalization of the area under the ROC curve (AUC). An AUC or C-statistic of 1.0

indicates perfect predictive discrimination, and an AUC or C-statistic of 0.50 indicates that a test that does not discriminate between overcrowding and no overcrowding. An AUC or C-statistic of at least 0.80 is considered to have good discrimination.[8] Because a model based on data obtainable from one of the scales would simplify data collection, we compared the performance of each model with a combined model consisting of all of the variables used in both scales. The comparison was made by using likelihood ratio (LR) chi-square statistics, which are sensitive measures of model fit. A measure of adequacy was determined on the basis of the ratio of the LR statistics with a scale alone, compared with the overall model likelihood ratio statistics. This is a unitless index of adequacy of a subset of predictors, or here, of the individual scales. To compare the superiority of NEDOCS over EDWIN and vice versa, a single-factor chi-square test was used by comparing the LR chi-square statistics of the models with each scale alone.[9] This LR model is well cited statistically and has been used elsewhere to compare study measurements.[10,11]

RESULTS

There were 131 sampling times over ten days. Only one sampling time was missed during that period, for a total of 130 completed sampling times. The VAS range for the outcome variable was 4–91 mm with a mean (±SD) of 51 (±24.6). The median VAS was 56, and the interquartile range (IQR) was 29 to 72 mm. The OV indicated that the ED was overcrowded in 80 (62%) of the 130 sampling times.

The median score for the NEDOCS was 93 (IQR: 72, 112), with a mean (±SD) score of 91 (±28.8) and a range of 31 to 144. The median score for EDWIN was 1.54 (IQR: 1.33, 1.83), the mean score was 1.58 (±0.43), and the range was 0.60 to 2.62.

The Pearson correlation coefficient for testing the association between NEDOCS and OV was 0.71, and for EDWIN and OV it was 0.74 (both, $p < 0.001$). EDWIN and NEDOCS results are highly correlated with each other and with the expert opinion on overcrowding, as shown in Figure 1 ($r = 0.84$; $p < 0.001$).

When the EDWIN and the NEDOCS were compared with the overcrowding variable, ROC curves were reasonably similar (Figure 2). The AUC for the NEDOCS was 0.83 (95% CI = 0.75 to 0.90), and the AUC for the EDWIN was 0.80 (95% CI = 0.73 to 0.88). Cutoffs were determined for both scales on the basis of a sensitivity of 80%. A cutoff of 87 on the NEDOCS was 80% sensitive and 71% specific, whereas a cutoff of 1.40 on the EDWIN was 80% sensitive and 63% specific for overcrowding.

The C-statistic for the logistic regression model for NEDOCS was 0.83. The IQR odds ratio, comparing the 75th percentile with the 25th percentile (112 vs. 73 NEDOCS), was 8.61 (95% CI = 3.99 to 18.54). The C-statistic for the logistic regression model for EDWIN was 0.81. The IQR odds ratio, comparing the 75th percentile with the 25th percentile (1.83 vs. 1.33 EDWIN), was 6.61 (95% CI = 3.17 to 13.79).

The LR chi-square statistics for NEDOCS and EDWIN alone, and a combined model containing both the NEDOCS and EDWIN, are given in Table 1. For EDWIN, the adequacy with respect to predicting overcrowding, compared with the combined model, was 86% (43.13/50.14). For NEDOCS, the adequacy was 97% (48.56/50.04). NEDOCS was found to add significantly more than EDWIN to the combined model (χ^2 test = 5.43; $p = 0.02$).

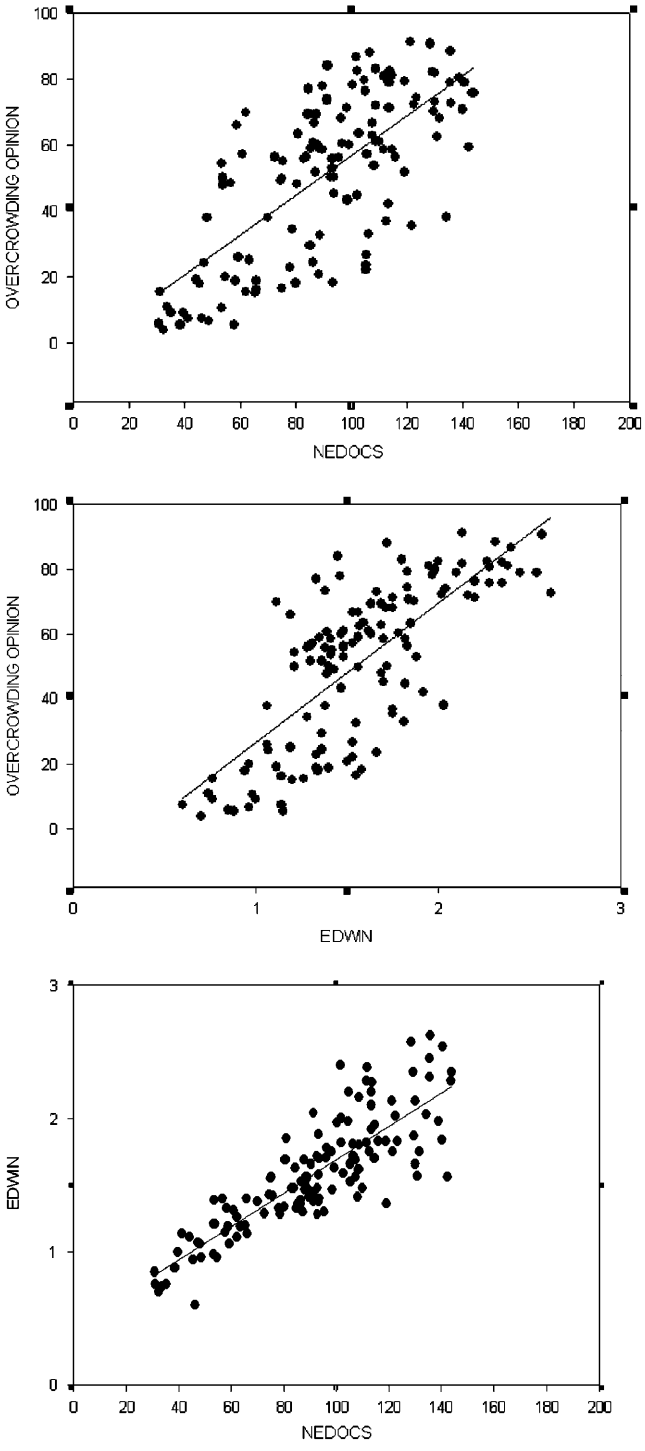

Figure 1. Comparison of results of the Emergency Department Work Index (EDWIN), the National Emergency Department Overcrowding Scale (NEDOCS), and overcrowding expert opinion. Correlation coefficients (*r*) for the comparison are 0.71 (*top*), 0.75 (*middle*), and 0.84 (*bottom*).

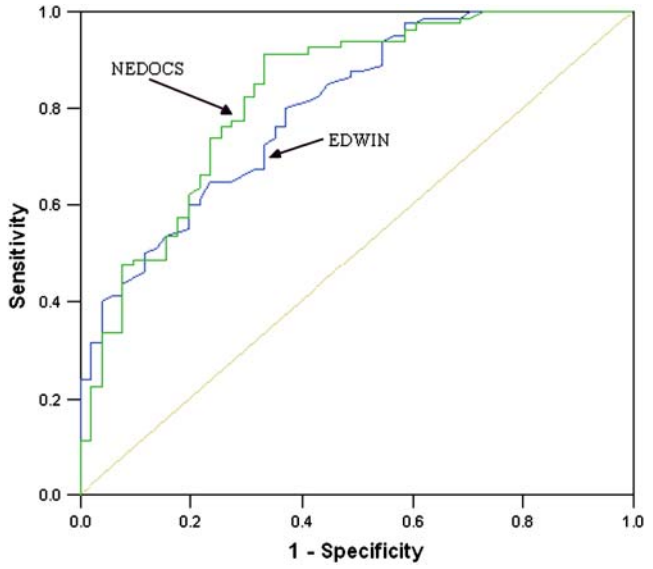

Figure 2. Receiver–operator curves for the National Emergency Department Overcrowding Scale (NEDOCS) and the Emergency Department Work Index (EDWIN). Area under the curve is 0.83 for the NEDOCS and 0.80 for the EDWIN.

DISCUSSION

Emergency department overcrowding has been related to multiple adverse outcomes. The patient's length of stay increases, thus increasing the number of patients who leave the ED without being seen,[12–15] some of whom may have diseases that are as much of a significant medical emergency as those of patients who stay to be seen.[12,13,15–21] Among those who stay to be seen, ED overcrowding affects quality of care[22–25] and patient satisfaction.[26–31] Inadequate patient care leads to medical errors as an ED becomes more overcrowded.[32–36] Some of these errors can be severe enough to lead to death and disability.[37] For all of these reasons, EDs must continue to work through administrative and political channels to prevent overcrowding.

In this study, we have attempted to improve our understanding of the construct of overcrowding. Not only are both scales evaluated here well correlated to a standardized ED overcrowding outcome variable, but they also are well correlated with each other. This study acts both as a prospective validation of the two scales and as an affirmation of their construct validity. The NEDOCS and EDWIN scales also had high AUCs and showed excellent sensitivity and specificity for ED overcrowding.

This study demonstrates a superiority of the NEDOCS when compared with the EDWIN in measuring overcrowding. The difference between the scales is relatively small, and often conditions in the ED will dictate which of these scales is a better choice. For example, the EDWIN is based on the ESI; if the ESI triage scoring system is not used in a particular ED, an EDWIN score cannot be calculated. Alternatively, if times from registration and from admission decisions are hard to obtain, the NEDOCS score cannot be calculated. Each ED must therefore determine which scale can be best applied.

Table 1
Adequacy* of the NEDOCS and EDWIN

Scales	Chi-Square	Adequacy (%)
EDWIN	43.13	86
NEDOCS	48.56	97
Combined	50.04	100

* Adequacy in reflecting the results of a model consisting of all of the variables used in each of the two models, combined into a single model.

The NEDOCS and the EDWIN both have face validity. The NEDOCS was designed on the basis of expert input from eight ED sites nationwide and was developed statistically by reducing a 20-question model to the best 5 questions. Face validity of the EDWIN is based on an intuitive understanding of ED overcrowding. Both scales have helped to advance the science of overcrowding. Research can now begin to evaluate the effect of ED systems changes on adverse outcomes by using the quantitative variables reflecting overcrowding.

LIMITATIONS

Emergency department overcrowding research is limited by the problem of a soft criterion standard. Both overcrowding scales studied here were designed and validated on the basis of the standard of emergency providers' opinion of overcrowding.[3,5] Although not the perfect criterion standard, ED provider opinion appears to be a consistent marker for the construct of overcrowding. It is as clear as many other criterion standards that have been used in other research areas of emergency medicine, such as in pulmonary embolism studies, in which the criterion standard constantly is changing.[38–40] ED expert opinion is the only logical starting point for development of quantitative measures of overcrowding.

Using a VAS to determine expert opinion has some drawbacks. Overcrowding is not necessarily a continuous phenomenon. Both NEDOCS and EDWIN were derived by using Likert-like scales. The use of a continuous scale, with markings in proximity to the line, appeared more appropriate to fairly reflect overcrowding and allow comparison of the two scales.

This study was performed at a single academic ED. An important factor in the applicability of our results is that neither the NEDOCS nor the EDWIN were derived in our ED. Therefore, we believe that our results can be generalized to other EDs for both of the scales, thus increasing the validity of the results.

Another limitation is that constructs used to model complex behavior must be tested in almost every possible type of setting before they are determined to be generalizable and valid. Neither of these two scales has yet been tested that extensively. The NEDOCS has been validated in numerous busy academic centers, whereas the EDWIN has been validated in one academic center. This study represents one further step in the process of validating these scales.

CONCLUSIONS

Both scales had high AUCs, correlated well with each other, and showed good discrimination for predicting ED overcrowding. This establishes construct validity for

these scales as measures of overcrowding. Which scale is used in an ED is dependent on which set of data is most readily available, with the favored scale being the NEDOCS.

References

1. Hwang U, Concato J. Care in the emergency department: how crowded is overcrowded? Acad Emerg Med. 2004; 11:1097–101.
2. Weiss SJ, Arndahl J, Ernst AA, Derlet R, Richards J, Nick TG. Development of a site sampling form for evaluation of ED overcrowding. Med Sci Monit. 2002; 8:CR549–53.
3. Bernstein SL, Verghese V, Leung W, Lunney AT, Perez I. Development and validation of a new index to measure emergency department crowding. Acad Emerg Med. 2003; 10:938–42.
4. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CA Jr. A conceptual model of emergency department crowding. Ann Emerg Med. 2003; 42: 173–80.
5. Weiss SJ, Derlet R, Arndahl J, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS). Acad Emerg Med. 2004; 11:38–50.
6. Weiss SJ, Ernst AA, Derlet R, King R, Bair A, Nick TG. Relationship between the National ED Overcrowding Scale and the number of patients who leave without being seen in an academic ED. Am J Emerg Med. 2005; 23(3):288–94.
7. Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The emergency severity index triage algorithm version 2 is reliable and valid. Acad Emerg Med. 2003; 10:1070–80.
8. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36.
9. Harrell FE. Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression and Survival Analysis. New York: Springer-Verlag, 2001.
10. Califf RM, Phillips HR 3rd, Hindman MC, et al. Prognostic value of a coronary artery jeopardy score. J Am Coll Cardiol. 1985; 5:1055–63.
11. Ridker PM, Rifai N, Clearfield M, et al. Measurement of C-reactive protein for the targeting of Statin therapy in the primary prevention of acute coronary events. N Engl J Med. 2001; 344:1959–65.
12. Davis B, Sullivan S, Levine A, Dallara J. Factors affecting ED length-of-stay in surgical critical care patients. Am J Emerg Med. 1995; 13(5):495–500.
13. Bucheli B, Martina B. Reduced length of stay in medical emergency department patients: a prospective controlled study on emergency physician staffing. Eur J Emerg Med. 2004; 11:29–34.
14. Fernandes CM, Price A, Christenson JM. Does reduced length of stay decrease the number of emergency department patients who leave without seeing a physician? J Emerg Med. 1997; 15:397–9.
15. Partovi SN, Nelson BK, Bryan ED, Walsh MJ. Faculty triage shortens emergency department length of stay. Acad Emerg Med. 2001; 8:990–5.
16. Baker DW, Stevens CD, Brook RH. Patients who leave a public hospital emergency department without being seen by a physician. Causes and consequences. JAMA. 1991; 266:1085–90.
17. Stock LM, Bradley GE, Lewis RJ, Baker DW, Sipsey J, Stevens CD. Patients who leave emergency departments without being seen by a physician: magnitude of the problem in Los Angeles County. Ann Emerg Med. 1994; 23:294–8.
18. Weissberg MP, Heitner M, Lowenstein SR, Keefer G. Patients who leave without being seen. Ann Emerg Med. 1986; 15:813–7.
19. Hobbs D, Kunzman SC, Tandberg D, Sklar D. Hospital factors associated with emergency center patients leaving without being seen. Am J Emerg Med. 2000; 18:767–72.
20. McMullan JT, Veser FH. Emergency department volume and acuity as factors in patients leaving without treatment. South Med J. 2004; 97:729–33.
21. Polevoi SK, Quinn JV, Kramer NR. Factors associated with patients who leave without being seen. Acad Emerg Med. 2005; 12:232–6.
22. Miro O, Antonio MT, Jimenez S, et al. Decreased health care quality associated with emergency department overcrowding. Eur J Emerg Med. 1999; 6:105–7.
23. Lombrail P, Vitoux-Brot C, Bourrillon A, Brodin M, De Pouvourville G. Another look at emergency room overcrowding: accessibility of the health services and quality of care. Int J Qual Health Care. 1997; 9:225–35.
24. Hagland MM. ED overcrowding spurs interest in quality and credentialing issues. Hospitals. 1991; 65:33–6.
25. Lindsay P, Schull M, Bronskill S, Anderson G. The development of indicators to measure the quality of clinical care in emergency departments following a modified-delphi approach. Acad Emerg Med. 2002; 9:1131–9.
26. DeBehnke D, Decker MC. The effects of a physician-nurse patient care team on patient satisfaction in an academic ED. Am J Emerg Med. 2002; 20(4):267–70.
27. Thompson DA, Yarnold PR, Williams DR, Adams SL. Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. Ann Emerg Med. 1996; 28:657–65.
28. Nerney MP, Chin MH, Jin L, et al. Factors associated with older patients' satisfaction with care in an inner-city emergency department. Ann Emerg Med. 2001; 38:140–5.
29. Boudreaux ED, D'Autremont S, Wood K, Jones GN. Predictors of emergency department patient satisfaction: stability over 17 months. Acad Emerg Med. 2004; 11:51–8.
30. Boudreaux ED, Mandry CV, Wood K. Patient satisfaction data as a quality indicator: a tale of two emergency departments. Acad Emerg Med. 2003; 10:261–8.
31. Boudreaux ED, Ary R, Mandry C. Emergency department personnel accuracy at estimating patient satisfaction. J Emerg Med. 2000; 19(2):107–12.
32. Anonymous. Disclosure of medical errors. Ann Emerg Med. 2004; 43:432.
33. Fordyce J, Blank FS, Pekow P, et al. Errors in a busy emergency department. Ann Emerg Med. 2003; 42: 324–33.

34. Hobgood C, Xie J, Weiner B, Hooker J. Error identification, disclosure, and reporting: practice patterns of three emergency medicine provider types. Acad Emerg Med. 2004; 11:196–9.
35. Feldman JA. Medical errors and emergency medicine: will the difficult questions be asked, and answered? Acad Emerg Med. 2003; 10:910–1.
36. Luszczak M, Kharod CU, Abbrescia K. Medical mistakes, errors in judgment, and personal awareness. Ann Emerg Med. 2003; 41:155–6.
37. Derlet RW. Triage and ED overcrowding: two cases of unexpected outcome. Can J Emerg Med. 2002; 3:8–9.
38. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a "gold standard". Am J Psychiatry. 1994; 151:650–7.
39. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. Am J Epidemiol. 1993; 137:1251–8.
40. Richman PB. The paradox of gold standards. Acad Emerg Med. 2002; 9:710–2.